nature research

Corresponding author(s): Gerald Crabtree

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Using previously published data and pilot studies, we calculated the difference between the means of the null and observed distributions at given effect sizes. We then estimated the sample size necessary to reduce the s.e.m. to achieve a statistical power > 80%.

2. Data exclusions

Describe any data exclusions.

No data were excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experimental findings were reliably reproduced with biological replicates for all experiments.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Samples were randomly assigned to experimental conditions, to processing order and to positions on plates where applicable.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded to group allocations.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Bowtie2, MACS2, CQN, Samtools, BEDTools, MAVRIC, Prism (7), FlowJo, R.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions on availability of materials.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	All antibodies were validated in pilot experiments. CD45: Biolegend, 30-F11, Cat#103126; Ly51: Biolegend, 6C3, Cat#108308; MHC-II I-A/I-E: Biolegend, M5/114.15.2, Cat#107618; Epcam (CD326): Biolegend, G8.8, Cat#118216; UEA-I: Vector Laboratories, Cat#FL-1061 TCR-beta: Biolegend, H57-597, Cat#H57-597; CD4: Biolegend, RM4-5, Cat#100531; CD8: Biolegend, 53-6.7, Cat#100714; CD25: Biolegend, PC61, Cat#102012; CD69: Biolegend, H1.2F3, Cat#104508; CD44: Biolegend, IM7; Cat#103030; CD62L: Biolegend, MEL-14; Cat#104412; Aire: eBiosciences/Thermo, 5H12; Cat#53-5934-82; Foxp3: eBiosciences/Thermo, FJK-16s, Cat#12-5773-82; V5:Thermo Fisher,Cat#R960-25 CD3: DakoCytomation, Cat#A045229

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Mouse embryonic fibroblasts (MEFs) were acquired from mice with the CiA:Oct4 allele at embryonic day 14.5 as previously described33. MEFs were transformed with simian virus 40 large T antigen and single cell sorted after transfection with LGmCreER (self-deleting) plasmid53 (Addgene #33340) to enrich for cells with excised neo cassette. Clones were screened for growth rate, VP16-mediated eGFP activation and DNase accessibility at the CiA:Oct4 locus. A single clone was used for all CiA recruitment experiments. MEFs were grown in high-glucose DMEM (Life Technologies, 11960) supplemented with 10% FBS (Omega Scientific, FB-11), 10 mM HEPES pH 7.5, Minimal AA, glutaMAX, Na Pyruvate, Pen/Strep, 2-Mercaptoethanol at 37o C., 5% CO2. The 4D6 human thymic epithelial cell line37 was a gift from Maria Toribio and Diane Mathis. The 4D6 cells were grown in RPMI 1640 (Life Technologies, 21870092) supplemented with 10% FBS (Omega Scientific, FB-11) and Pen/Strep at 37o C., 5% CO2.
b. Describe the method of cell line authentication used.	Genotyping and RT-PCR.
c. Report whether the cell lines were tested for mycoplasma contamination.	Cell lines tested negative for mycoplasma contamination.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

B6. Aire+/- mice5 obtained from Jackson Laboratory were bred to generate Aire-/- and Aire+/+ littermates. Brg1F/F and Foxn1ex9cre mice51,52 on C57BL/6 background were bred to generate Brg1F/F and Foxn1ex9cre;Brg1F/F littermates. All experiments used female and male littermate controls. All mice were maintained in accordance with Stanford University's Animal Care and Use Committee guidelines.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Study did not involve human research participants.

nature research

Corresponding author(s): Gerald Crabtree

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Thymi were dissected, capsules incised and triturated with glass pipettes to release thymocytes from stromal fragments. Stroma were digested with Liberase TM (Roche) and DNase I (Roche), and TECs were enriched using α-CD45 MACS microbeads (Miltenyi) or centrifugation on a Percoll PLUS gradient (GE Healthcare). Cells were treated with Fc-block, stained on ice for 20 min, washed and analyzed.
6. Identify the instrument used for data collection.	FACS Aria II and LSR II flow cytometer (BD).
7. Describe the software used to collect and analyze the flow cytometry data.	FlowJo
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Post-sort fractions were analyzed to have > 95% of the relevant cell population.
9. Describe the gating strategy used.	See Supplementary Figure 1a.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒

nature research

Corresponding author(s): Gerald Crabtree

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. *The entry may remain private before publication.*	See Data Availability section in Methods.
3. Provide a list of all files available in the database submission.	Raw SRA files, Processed count matrices.
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).	Can provide upon request.

Methodological details

5. Describe the experimental replicates.	Two biological replicates for each cell type.
6. Describe the sequencing depth for each experiment.	50bp paired-end reads. Unique reads for each sample: 20687490 WTHIrep1 42056204 WTHIrep2 33850722 AireHIrep1 31685034 AireHIrep2 53535020 WTLOrep1 46310276 WTLOrep2 51667712 AireLOrep1 43074392 AireLOrep2 30550828 NOCREHIrep1 25840370 NOCREHIrep2 41213098 BRGHIrep1 37725772 BRGHIrep2 42358570 NOCRELOrep1 29237962 NOCRELOrep2 40336154 BRGLOrep1 41880940 BRGLOrep2
7. Describe the antibodies used for the ChIP-seq experiments.	ATAC-seq Tn5 from Illumina Nextera Library Kit.
8. Describe the peak calling parameters.	See ATAC-seq data analysis section in Methods.
9. Describe the methods used to ensure data quality.	See ATAC-seq data analysis section in Methods.
10. Describe the software used to collect and analyze the ChIP-seq data.	Bowtie2, MACS2, CQN, Samtools, BEDTools, MAVRIC, R.